Introduction to Data Integration Driven by a Common Data Model

Michal Džmuráň
Senior Business Consultant
Progress Software Czech Republic

Index

Data Integration Driven by a Common Data Model

Data Integration Engine

Today, not even small organisations can make do with a single application. The majority of business processes in an organisation is nowadays already supported by some kind of implemented application, and the organisation must focus on making its operations more efficient. One of the ways of achieving this goal is the information exchange optimization across applications, assisted by the data integration from various applications; it is summarily known as Enterprise Application Integration (EAI). Without effective Enterprise Application Integration, a modern organisation cannot run its business processes to meet the ever increasing customer demands and have an up-to-date knowledge of its operations. Data integration is an important part of EAI. Since SOA (Services Oriented Architecture) plays today more and more important role in EAI, data integration has an unevitable role in it. A few independent resources emphasize the importance of the data integration in the SOA context. Let's bring out two. According to Gartner[1], it is the data integrity which drives the complexity of SOA solutions and why SOA does not achieve the expectations and goals expected. Martin Nečaský writes about data integration[2]: "Successful data integration is an important, yet very difficult challenge with many information systems. The data integration issue can generally be described as having a number of Data Sources on one side, for storing data. On the other side, there is a bunch of Data Services that use the Data Sources and support business process segments in an enterprise.

Each such service accesses the data from one or more sources, processes it and sends it to its clients in a standard format defined in its schema. We can assume that Data Services send data to clients in the form of XML documents. A Data Source may provide data generally in any representation described in the schema. It would typically be a database (object-oriented, relational, XML), XML data source (e.g. a web service) or another API interface, etc.

Data Sources and Services are usually independent of each other. As a result they have their own structures and/or schemas which may be very different."

What is Semantic Integration?

Semantic Integration is the part of the Enterprise Application Integration (EAI), or more specifically of data integration, that focuses on data exchange across applications based on their semantics, content and the required business rules. Examples:

- The name "John Doe" is syntactically different from "Doe John"; however semantically, these two names are the same provided they relate to the same person. Semantic Integration detects such matches and treats both variations of the same name as one person.

- The insurance policy number "123456" is syntactically always the same, but in reality it may refer to two policies – one life insurance policy and the other property insurance. The semantic integration role is to detect the difference and always choose the right policy to work with.
- Similar data (from the point of defined business rules) is stored in different syntactic formats in different databases. Business rules however demand that the similar data be presented in a standardised format, which also is not the same as any of the formats used in the database. Semantic Integration consolidates the data and transforms it into the required format.

What Is Common Data Model? Common Data Model Examples

A Common Data Model (or sometimes referred to as Canonical Data Model, or Common Model in short) is an application-independent data model describing the structure and data semantics in relation to the organisation's business processes. It is typically created using one of the modelling tools and described using a modelling language – UML being most commonly used today. Generally speaking, a Common Data Model should contain descriptions of all the data used by the organization. Such a Common Data Model could be very extensive and its benefits set against the cost of its making would be debatable. That is why Common Data Models often include only data which is integrated by applications. A proved practice would be to use the physical data model of the dominant data source as the basis for the Common Data Model. Some organisations form associations to develop a Common Data Model for their industry. Such models stand a good chance to become a generally accepted standard in the industry in question. Some Common Models are created from the government initiative. Such models are usually too extensive to be used by a single organisation. Organisations working with the models adopt only the parts relevant to them. Common Data Model examples taken as industry standards are shown at the end of **Getting Started with the SID**[3] by John Reilly.

What Is Data Integration Driven by a Common Data Model?

Data Integration Driven by a Common Data Model is a technological Enterprise Application Integration concept using the Common Data Model as a common layer for the data exchange. According to research by Forrester for Progress Software, the tool that enables such data integration is the tool-of-choice for data sharing and data integration with the company's services – see Chart 1. A Common Data Model serves as a backbone system with whom individual Data Sources as general information sources – and Data Services as information consumers – communicate. Data Integration does not address the definition and the Business Process Management. As such, however, it is a success factor in its implementation.

Data Integration Driven by a Common Data Model

As mentioned above, the Common Data Model describes the data semantics. Data Integration Driven by a Common Data Model therefore has all the qualities of Semantic Data Integration. The benefits of this type of integration include:

- Elimination of the so-called spaghetti integration when each application communicates with every other one.
- Reduction in the transformation number required to convert M×N to M+N when each application talks to every other one, with only transformations of each application to the Common Data Model developed.
- Enterprise data integration from all sources in the common structure, with a preference for semantics over syntax.
- Simpler implementation of future changes. Some changes in master applications would be implemented in one location on the level of the Common Data Model with transformation.

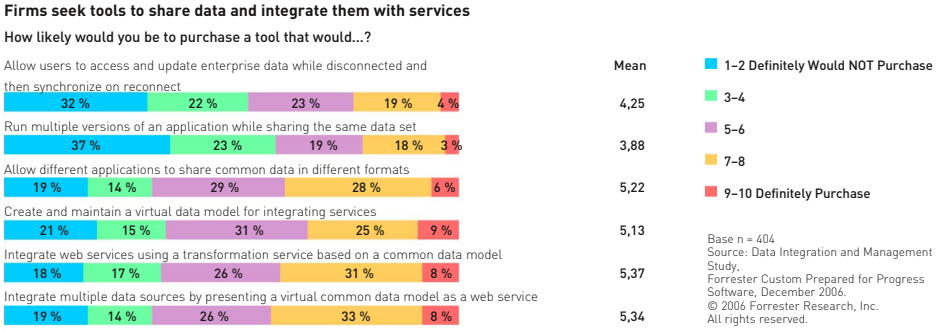

Chart 1: Tool preference for data sharing and data integration with services

Figure 1 illustrates the differences in using the point-to-point arrangement vs. the Common Data Model for the data integration.

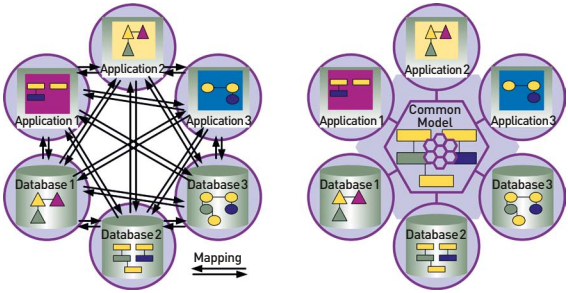

Figure 1: Point-to-Point vs. Common Data Model

Data Integration Driven by a Common Data Model

Position of the Data Integration Driven by a Common Data Model in the Application Integration Architecture

Gartner defined four possible approaches to application integration:

- Message Oriented Middleware (MOM);
- Application server;
- Comprehensive package solution; and
- Enterprise Services Bus (ESB).

Some sources, e.g.[4], also mention file transfer integration, but modern practice already regards this approach as obsolete. Then how is it possible to implement Data Integration Driven by a Common Data Model in the individual approaches?

- In the SOA era, Message Oriented Middleware (MOM) is useful especially where high performance in real time is required – thousands of messages forwarded between a limited number of systems, without the need for complex transformations. The benefit of Data Integration Driven by a Common Data Model is debatable; moreover, it could reduce the performance below an acceptable level.
- The application server is a good choice for this type of integration. The market knows many types of application servers that can be implemented for this kind of data integration by means of Web Services or EJB. Implementing this kind of integration in the application server environment has one significant advantage – the spaghetti integration elimination to which this environment is otherwise quite .
- Implementing this kind of data integration in a comprehensive package solution must always be an individual project; it cannot be adequately described in general terms.
- Enterprise Service Bus (ESB) is another good candidate for an environment where Data Integration Driven by a Common Data Model can be easily implemented, in this instance in the form of services connected to the ESB.

The ESB technological concept even overlaps the concept of Data Integration Driven by a Common Data Model. The schema and data transformation issues and routing based on the message content are addressed in the same way by data integration and ESB services. In practical terms, it has proven to be optimal to leave process functions on the ESB level, and address the rest by data integration (see[5]).

The ESB environment offers various options for data integration implementation. The implementation method largely depends on the individual product capacity available for collaboration.

Data Integration Driven by a Common Data Model

In the first instance, a data integration adapter can be implemented between the ESB and the application in question. The data integration itself is not done on the ESB level, but in the end systems. It may however require modifications in the end systems, which is a possible drawback of this solution. Figure 2 illustrates the architecture after the data integration layer is implemented between the ESB and the Data Source, and between the ESB and the Data Service.

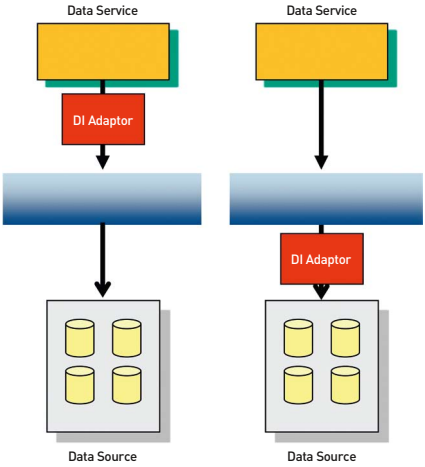

Figure 2: Data integration as adapter to ESB

In the second instance, the data integration is implemented using Web Services running in the Java application server environment connected to the ESB. Figure 3 shows this example. This method is non-invasive with respect to the end application and data sources.

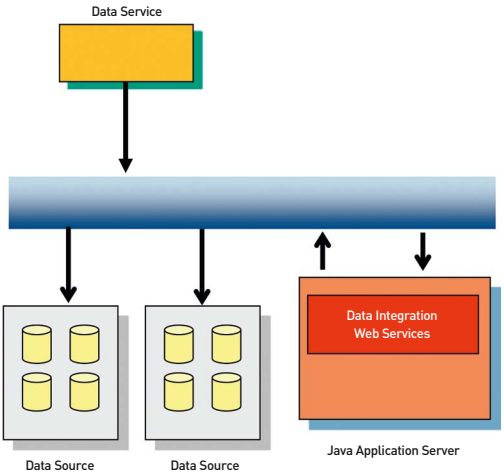

Figure 3: Data integration as web service to ESB

Data Integration Driven by a Common Data Model

Figure 4 illustrates the third option – the data integration implementation directly into the ESB context. This option is also non-invasive to the Data Sources and Services. Today we have only a limited range of ESB and data integration products capable of such collaboration. The different data integration implementation method in the ESB environments are not mutually exclusive and may be combined in a specific project.

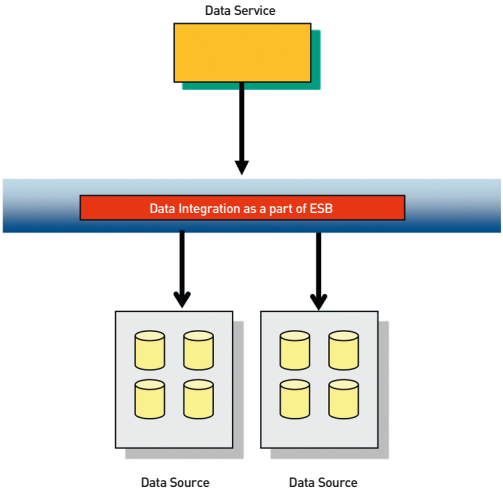

Figure 4: Data integration as a part of ESB

Which Organisations Would Benefit from Data Integration Driven by a Common Data Model?

The technological concept of Data Integration Driven by a Common Data Model is suitable especially for organisations with many data sources (also referred to as Operating System Services, OSS) and many business processes (also referred to as Business System Services, BSS) using the data from the multiple sources. Such structures would typically develop as the organisation grows, but may occur also as a result of a merger, which are quite frequent in these globalized times.

Examples:

- Telecommunications operators managing fixed and mobile telephone networks and offering also internet services. Every such activity typically originates in a separate system.
- Multinational companies needing a comprehensive overview of their operations, requiring data from individual national systems in real time.
- Trading companies with multiple customer classes (retail, wholesale). The customer data is stored in different systems, but the companies wish to provide its services via a single portal.

Key Elements of Data Integration Driven by a Common Data Model

Common Data Model, Data Services and Data Sources

The Common Data Model is a common data integration denominator which it manages. The model is in fact a backbone layer used to exchange data between Data Services and Data Sources.

Data Services are applications that read and transform data. Data Services have their own schemas describing how the data is presented, and defined operations for the query initiation into the Common Data Model. In the data integration context, a Data Service may be the Data Source.

A Data Source is an application, database or another data repository. Data Sources have their own schemas. They do not initiate any operations themselves; they are approached by the Data Services via the Common Data Model. A Data Source can be a database, but also a Web Service, HTTP query, etc. In the data integration context, a Data Source can also be the Data Service.

Figure 5 schematically illustrates the Domain Model as an aggregate of the Common Data Model, Data Services and Data Sources. The arrows indicate transformations.

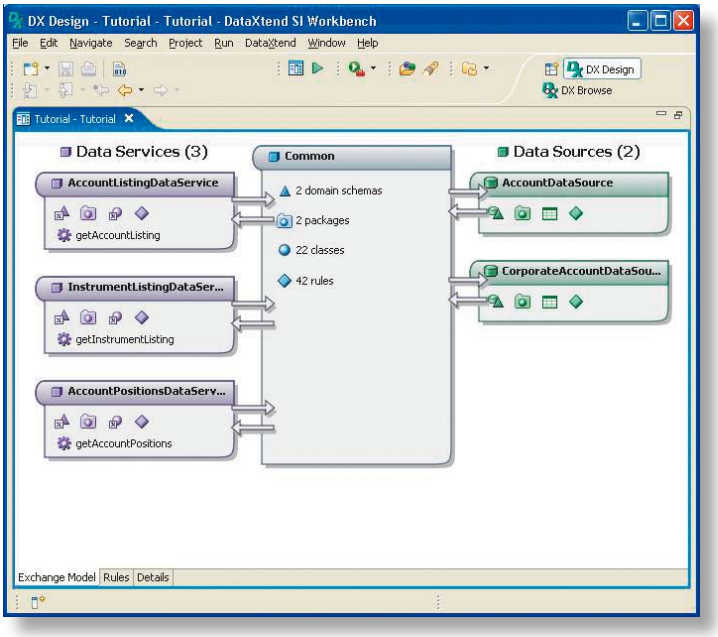

Figure 5: Domain Model

Key Elements of Data Integration Driven by a Common Data Model

Data transformations, which are necessary with a view to their different syntax in different Data Services and Data Sources, rules for aggregation of the same data from different Data Sources and validation rules for Data Service requirements are implemented on the Common Data Model level.

Mapping and Computed Attributes

Mapping is a system of defined pairs which determine how the Data Service and Data Source classes and attributes are implemented in the Common Data Model and vice versa. Mapping is generally only one-way – if, for instance, an account number in the Common Data Model is mapped to the account number in the Data Source (by a query to the Data Source), it does not mean that the reply (results returned by the Data Source) need to map the account number to the same class and attribute in the Common Data Model.

Figure 6 illustrates the Data Source or Data Service mapping to the Common Data Model. The one-way arrows indicate the mapping process as a one-way process. Computed attributes are an added part of the Common Data Model that implements the data semantics. They are used for the transformation rule definition which take into account the different data syntax in the applications included in the Common Data Model[1].

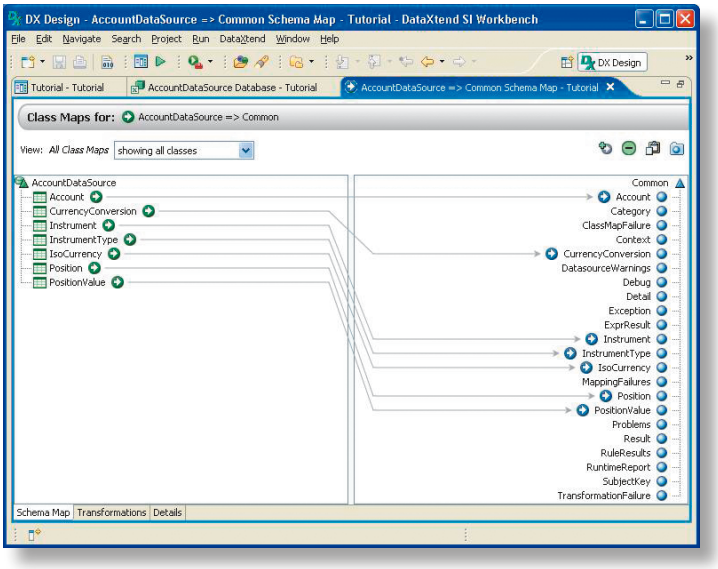

Figure 6: Attribute mapping

Key Elements of Data Integration Driven by a Common Data Model

Rules

Rules are implemented business constraints defined on the Common Schema, Data Source or Data Service level.
The rules on the Data Source and Data Service level are based on individual application constraints, e.g. the database dictionary. The rules need not necessarily be included in the data integration, but often are for performance reasons. The rules include, for instance, the primary key uniqueness rule, etc.
The rules on the Common Data Model level are based on the data integration needs and derived from business processes defined across the applications. They include, for instance, a constraint on the validity of values entered in orders online.

Example:
The following simple example illustrates how data integration works in practice. For the sake of keeping it simple, the example does not include the full number of Data Sources and Data Services.

A telecommunications operator has three main commodities. The company operates a fixed line network, a mobile network and provides internet services. There are the following Data Sources which originate in the history:

- **Fixed line network**
  this data source grew as the organisation developed. It has its own customer database, a database of numbers, tariff plans, connected calls, etc.
- **Mobile network**
  this data source is a legacy after the acquisition of a mobile operator. As such, it also has its own customer database, a database of mobile numbers, pre-paid SIM cards, tariff plans, connected calls, etc.
- **Internet services**
  this data source is also a legacy after the acquisition of an ISP. As such, it also has its own customer database, a database of service bundles subscribed, etc.

The databases are vastly different – both in terms of their model and their syntax. Sometimes a full name is handled as a single attribute, sometimes as two separate values. The telephone number formats are different, as are the methods of assigning numbers to customers, etc.

Key Elements of Data Integration Driven by a Common Data Model

The company wants to implement the following business processes to its data services provision:

- **A single point of service, central billing**
  A single point of service is a service where the customer uses more services from the main commodity areas and which provides services to the customer that span all these commodities. These services may be, for instance, special price bundles or central billing – the customer gets a regular total bill for all the services provided from all the main commodities.
- **Overview of the customer's services**
  This query to the data sources would return a full picture of all services subscribed by the customer in question. The overview contains information about fixed line and mobile services, as well as about internet services.
- **Service analysis**
  A service analysis is a process used for improving the organisation's service quality. Its purpose is to find customer(s) who use(s) services from multiple commodities but are served separately for each commodity in various data sources, using matches or similarities of selected attributes. The organisation wants to identify the customers so that it can migrate them to a single point of service.

The organisation may choose, for instance, the original data model used by the fixed line network system as its Common Data Model basis, and expand it by the requisite data from the other systems; or it may use the industry standard SID Common Data Model for telecommunications (**www.tmforum.org**).

Glossary of Terms

Aggregation Rule
A rule for resolving duplicate entities with the same syntax but different semantics from different data sources on the same Common Data Model level, or for the aggregation of syntactically different, but semantically corresponding entities from different data sources.

Business Integrity
The data validity based upon its semantic value in a business process. Data can generally meet integrity requirements or have validity in a specific application, but it is not necessarily semantically valid in the context of a business process spanning several Data Sources.

Class Map
A specification of correspondences between classes between the Common Data Model on the one hand and Data Sources and Services on the other, and vice versa. A class map is generally only a oneway map.

Common Data Model
A Common Data Model is an application-independent data model defining the data structure and semantics in relation to the business processes in an enterprise.

Computed Attribute
A computed attribute is defined on the Common Data Model level; it is used to implement syntactically different but semantically corresponding attributes on the Common Data Model level – in different Data Sources and Data Services.

Data Service
A Data Service is an application that reads and transforms data. A Data Service contains at least one operation, which initiates a query into the Common Data Model. In the data integration context, a Data Service can at the same time be the Data Source.

Data Source
A Data Source is an application, database or another data repository. In the data integration context, a Data Source can also be the Data Service. A Data Source typically serves up data in the XML format or via a JDBC interface.

Glossary of Terms

Domain Model
A Domain Model (referred also as Domain Schema) is a data model of the highest rank, incorporating the Common Data Model, Data Sources and Data Services. A Common Data Model is often based on an industry standard.

Operation
An operation is an externally visible procedure implemented in a Data Service, which itself implements a business process. An operation may include input parameters (classes) and return values (classes). An operation set constitutes the external data integration interface.

Rule
A predicate, constraint or preference used to validate individual operations or data. There are different rules – aggregation rule, transformation rule and validation rule.

Semantics
Semantics means looking at the real data meaning. Unlike the syntax which describes the formal data structure and format, semantics operates on the information merit.

Transformation Rule
A rule for deriving syntactically and semantically correct attributes of the target class from the source class attributes.

Validation Rule
A rule for ensuring that data meets semantic (consistency) requirements for data accuracy contained in a request from a Data Service.

Information Sources

Web

- **www.pantero.com** – Progress DataXtend SI website
- **http://www.progress.com/progress_software/dataxtend/docs/dataxtend_si_ds.pdf** – datasheet with selected general definitions
- **http://www.progress.com/dataxtend/data_integration/glossary_of_terms/index.ssp** – glossary of terms
- **http://www.billingworld.com/articles/feature/SOA-Success-Depends-on-Flexible-Data.html** – an article by John Petrie about the data integration importance
- **http://en.wikipedia.org/wiki/Semantic_translation** – Wikipedia

Articles and Analytical Reports

- **Datová integrace řízená společným schématem**, Martin Nečaský, Jindřich Štumpf, Progrese, autumn 2007
- **Progress DataXtend SI**, Martin Nečaský, Professional Computing, October 2007
- **Semantic Web Grounds Data Governance in World of Standardized Ontologies**, Jim Kobielus, Current Analysis, **http://www.currentanalysis.com/CCNET/CIR_25095.aspx** (paid link), August 2007.

Links to Common Data Models:

The links list is far from exhaustive. If you have any question relating to another Common Data Model, please contact the author at mdz@progress.com.

- **www.acord.org** – ACORD, insurance
- **www.tmforum.org** – SID, telecommunications
- **www.dmtf.org/standards/cim/** – CIM, (public) services
- **www.ppdm.org** (PPDM), **www.openm3.org** (MMDM) – utilities
- **www.openapplications.org** – OAGIS, manufacturing, supplier chains
- **www.hl7.org** (HL7), **www.hipaa.org** (HIPAA) – healthcare
- **www.nrf-arts.org** – (ARTS), sales
- www.fpml.org (FPML), **www.swift.com** (SWIFT) – capital markets

Literature:

- **Getting Started with the SID**: A SID Modeler's Guide, John P. Reilly, TeleManagement Forum, April 2007, ISBN: 978–0–9794281–0–4
- **Enterprise Service Bus**, David A, Chappell, O'Reilly Media, June 2004, ISBN: 978–0–596–00675–6

Information Sources

Contact:

Michal Džmuráň, Senior Business Consultant
Progress Software Czech Republic, Michelská 60/300, 140 00 Praha 4
Tel.: **+420 241 095 224**
Email: **mdz@progress.com**

References:

1 Presentation from Gartner Application Architecture, Development & Integration Summit, David Newman "Data Integration: At the Heart of the Data Services Evolution", June 11–13, 2007, Gaylord Opryland Resort & Convention Center, Nashville, TN

2 **Progress DataXtend SI**, Martin Nečaský, Professional Computing, October 2007

3 **Getting Started with the SID**: A SID Modeler's Guide, John P. Reilly, TeleManagement Forum, April 2007, ISBN: 978–0–9794281–0–4

4 **Technologie pro systémovou integraci**, Lenka Michalská, Computerworld CZ 11/2007

5 **Enterprise Service Bus**, David A. Chapell, O'Reilly Media, June 2004, ISBN: 978–0–596–00675–6

Notes

Michal Džmuráň is Senior Consultant for the Czech subsidiary of Progress Software, the leading supplier of data integration managed by a common schema. He has more than 20 years of experience in the area of development and operation of large-scale distributed information systems; he worked for Autoturist, Pragodata, ITC and presently for Progress Software. He often publishes articles on the current trends in the IT field and speaks at national and international conferences. Now he specializes in data integration managed by a Common Data Model and complex event processing.

PROGRESS
SOFTWARE

Prod code:
CZ08-04-06-013-Rev0